DOCUMENT RESUME

ED 435 714 TM 030 345

AUTHOR	Brualdi, Amy
TITLE	Traditional and Modern Concepts of Validity. ERIC/AE Digest.
INSTITUTION	ERIC Clearinghouse on Assessment and Evaluation, Washington, DC.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	EDO-TM-99-10
PUB DATE	1999-12-00
NOTE	4p.
CONTRACT	ED99CO0032
AVAILABLE FROM	ERIC Clearinghouse on Assessment and Evaluation, 1129 Shriver, University of Maryland, College Park, MD 20742. Tel: 800-464-3742 (Toll Free).
PUB TYPE	ERIC Publications (071) -- ERIC Digests in Full Text (073)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Criteria; Data Interpretation; Elementary Secondary Education; Reliability; *Test Interpretation; *Validity
IDENTIFIERS	ERIC Digests; *Messick (Samuel)

ABSTRACT

Test validity refers to the degree to which the inferences based on test scores are meaningful, useful, and appropriate. Thus, test validity is a characteristic of a test when it is administered to a particular population. This article introduces the modern concepts of validity advanced by S. Messick (1989, 1996, 1996). Traditionally, the means of accumulating validity have been grouped into the categories of content-related, criterion-related, and construct-related evidence of validity. Messick considered the traditional concept of validity to be fragmented and incomplete because it fails to take into account evidence of the value implications of score meaning as a basis for action and the social consequences of score use. Messick identified six aspects of validity that are general validity criteria or standards. These are: (1) content; (2) the substantive aspect; (3) structure; (4) generalizabilty; (5) external factors; and (6) consequential aspects of validity. Two major threats to validity that are worth noting are construct underrepresentation and construct-irrelevant variance. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 714

Traditional and Modern Concepts of Validity

Amy Brualdi

TM030345

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ERIC Digest EDO-TM-99-10

December 1999

BEST COPY AVAILABLE

ERIC®/AE Digest Series EDO-TM-99-10
University of Maryland, College Park

December 1999
College of Library and Information Services

Traditional and Modern Concepts of Validity

Amy Brualdi
Prince Williams Public Schools

Test validity refers to the degree with which the *inferences* based on test scores are meaningful, useful, and appropriate. Thus test validity is a characteristic of a test when it is administered to a particular population. Validating a test refers to accumulating empirical data and logical arguments to show that the inferences are indeed appropriate.

This article introduces the modern concepts of validity advanced by the late Samuel Messick (1989, 1996a, 1996b). We start with a briefly review the traditional methods of gathering validity evidence.

TRADITIONAL CONCEPT OF VALIDITY

Traditionally, the various means of accumulating validity evidence have been grouped into three categories -- content-related, criterion-related, and construct-related evidence of validity. These broad categories are a convenient way to organize and discuss validity evidence. There are no rigorous distinctions between them; they are not distrinct types of validity. Evidence normally identified with the criterion-related or content-related categories, for example, may also be relevant in the construct-related evidence

Criterion-related validity evidence - seeks to demonstrate that test scores are systematically related to one or more outcome criteria. In terms of an achievement test, for example, criterion-related validity may refer to the extent to which a test can be used to draw inferences regarding achievement. Empirical evidence in support of criterion-related validity may include a comparison of performance on the test against performance on outside criteria such as grades, class rank, other tests and teacher ratings.

Content-related validity evidence - refers to the extent to which the test questions represent the skills in the specified subject area. Content validity is often evaluated by examining the plan and procedures used in test construction. Did the test development procedure follow a rational approach that ensures appropriate content? Did the process ensure that the collection of items would represent appropriate skills?

Construct-related validity evidence - refers to the extent to which the test measures the "right" psychological constructs. Intelligence, self-esteem and creativity are examples of such psychological traits. Evidence in support of construct-related validity can take many forms. One approach is to demonstrate that the items within a measure are inter-related and therefore measure a single construct. Inter-item correlation and factor analysis are often used to demonstrate relationships among the items. Another approach is to demonstrate that the test behaves as one would expect a measure of the construct to behave. For example, one might expect a measure of creativity to show a greater correlation with a measure of artistic ability than with a measure of scholastic achievement.

MODERN CONCEPT OF VALIDITY

Messick (1989, 1996a) argues that the traditional conception of validity is fragmented and incomplete especially because it fails to take into account both evidence of the value implications of score meaning as a basis for action and the social consequences of score use. His modern approach views validity as a unified concept which places a heavier emphasis on how a test is used. Six distinguishable aspects of validity are highlighted as a means of addressing central issues implicit in the notion of validity as a unified concept. In effect, these six aspects conjointly function as general validity criteria or standards for all educational and psychological measurement. These six aspects must be viewed as interdependent and complementary forms of validity evidence and not viewed as separate and substitutable validity types.

- **Content** A key issue for the content aspect of validity is determining the knowledge, skills, and other attributes to be revealed by the assessment tasks. Content standards themselves should be relevant and representative of the construct domain. Increasing achievement levels or performance standards should reflect increases in complexity of the construct under scrutiny and not increasing sources of construct-irrelevant difficulty (Messick, 1996a).
- **Substansive** The substansive aspect of validity emphasizes the verification of the domain processes to be revealed in assessment tasks. These can be identified through the use of substansive theories and process modeling (Embretson, 1983; Messick 1989). When determining the substansiveness of test, one should consider two points. First, the assessment tasks must have the ability to provide an appropriate sampling of domain processes in addition to traditional coverage of domain content. Also, the engagement of these sampled in these assessment tasks must be confirmed by the accumulation of empirical evidence.

ERIC® ERIC Clearinghouse on Assessment and Evaluation, 1129 Shriver Laboratory,
University of Maryland, College Park, MD 20742 • 800 464-3742

- **Structure** Scoring models should be rationally consistent with what is known about the structural relations inherent in behavioral manifestations of the construct in question (Loevinger, 1957). The manner in which the execution of tasks are assessed and scored should be based on how the implicit processes of the respondent's actions combine dynamically to produce effects. Thus, the internal structure of the assessment should be consistent with what is known about the internal structure of the construct domain (Messick, 1989).
- **Generalizability** Assessments should provide representative coverage of the content and processes of the construct domain. This allows score interpretations to be broadly generalizable within the specified construct. Evidence of such generalizability depends on the tasks' degree of correlation with other tasks that also represent the construct or aspects of the construct.
- **External Factors** The external aspects of validity refers to the extent that the assessment scores' relationship with other measures and nonassessment behaviors reflect the expected high, low, and interactive relations implicit in the specified construct. Thus, the score interpretation is substantiated externally by appraising the degree to which empirical relationships are consistent with that meaning.
- **Consequential Aspects of Validity** It is important to accrue evidence of such positive consequences as well as evidence that adverse consequences are minimal. The consequential aspect of validity includes evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use. This type of investigation is especially important when it concerns adverse consequences for individuals and groups that are associated with bias in scoring and interpretation.

These six aspects of validity apply to all educational and psychological measurement; most score-based interpretations and action inferences either invoke these properties or assume them, explicitly or tacitly. The challenge in test validation, then, is to link these inferences to convergent evidence which support them as well as to discriminant evidence that discount plausible rival inferences.

SOURCES OF INVALIDITY

Two major threats to test validity are worth noting, especially with today's emphasis on high-stakes performance tests.

Construct underrepresentation indicates that the tasks which are measured in the assessment fail to include important dimensions or facets of the construct. Therefore, the test results are unlikely to reveal a student's true abilities within the construct which the test was indicated as having been measured.

Construct-irrelevant variance means that the test measures too many variables, many of which are irrelevant to the interpreted construct. This type of invalidity can take two forms, "construct-irrelevant easiness" and "construct-irrelevant difficulty." "Construct-irrelevant easiness" occurs when extraneous clues in item or task formats permit some individuals to respond correctly or appropriately in ways that are irrelevant to the construct being assessed; "construct-irrelevant difficulty" occurs when extraneous aspects of the task make the task irrelevantly difficult for some individuals or groups. While the first type of construct irrelevant variance causes one to score higher than one would under normal circumstances, the latter causes a notably lower score.

Because there is a relative dependence of task responses on the processes, strategies, and knowledge that are implicated in task performance, one should be able to identify through cognitive-process analysis the theoretical mechanisms underlying task performance (Embretson, 1983).

REFERENCES AND RECOMMENDED READING

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Embretson (Whitely), S. Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179-197.

Fredericksen, J.R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27-32.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635-694 (Monograph Supplement 9).

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Messick, S. (1996a). Standards-based score interpretation: Establishing valid grounds for valid inferences. *Proceedings of the joint conference on standard setting for large scale assessments*. Sponsored by National Assessment Governing Board and The National Center for Education Statistics. Washington, DC: Government Printing Office.

Messick, S. (1996b). Validity of Performance Assessment. In Philips, G. (1996). *Technical Issues in Large-Scale Performance Assessment*. Washington, DC: National Center for Educational Statistics.

Moss, P.A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*, 229-258.

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract ED99CO0032. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the U.S. Department of Education. ***Permission is granted to copy and distribute this ERIC/AE Digest.***